

# Bayesian Estimators in Uncertain Nested Error Regression Models

Shonosuke Sugasawa

*The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa-shi, Tokyo 190-8562, Japan.*

Tatsuya Kubokawa

*Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.*

---

## Abstract

Nested error regression models are useful tools for analysis of grouped data, especially in the case of small area estimation. This paper suggests a nested error regression model using uncertain random effects in which the random effect in each area is expressed as a mixture of a normal distribution and a positive mass at 0. For estimation of the model parameters and prediction of the random effects, an objective Bayesian inference is proposed by setting non-informative prior distributions on the model parameters. Under mild sufficient conditions, it is shown that the posterior distribution is proper and the posterior variances are finite, confirming the validity of posterior inference. To generate samples from the posterior distribution, we provide the Gibbs sampling method with familiar forms for all the full conditional distributions. This paper also addresses the problem of predicting finite population means, and a sampling-based method is suggested to tackle this issue. Finally, the proposed model is compared with the conventional nested error regression model through simulation and empirical studies.

*Keywords:* Bayesian estimator, nested error regression model, posterior propriety, small area estimation, uncertain random effect

---

## 1. Introduction

Linear mixed models and model-based estimators including the empirical Bayes estimator (EB) or empirical best linear unbiased predictor (EBLUP) have been studied quite extensively in the literature from both theoretical and applied points of view. Of these, small area estimation is an important application, and

---

*Email addresses:* shonosuke622@gmail.com (Shonosuke Sugasawa), tatsuya@e.u-tokyo.ac.jp (Tatsuya Kubokawa)

methods for small area estimation have received much attention in recent years due to a growing demand for reliable small area estimates. For good reviews on this topic, see Ghosh and Rao [8], Rao and Molina [13], Datta and Ghosh [6] and Pfeiffermann [12]. The linear mixed models used for small area estimation are categorized into two major types, the Fay-Herriot model suggested by Fay and Herriot [7] for area-level data, and the nested error regression (NER) models given in Battese, Harter and Fuller [2] for unit-level data. The resulting model-based estimators, such as EB or EBLUP, for small-cluster means or subject-specific values, provide reliable estimates with higher precision than direct estimates like sample means. These stable inferences are owing to random effects, but the misspecification of random effects may increase the risk of prediction.

Concerning this issue, Datta, Hall and Mandal [5] recently suggested inference by testing the presence of random effects in general mixed models. They pointed out that if the random effects can be dispensed, the model parameters and the small area means may be estimated with substantially higher accuracy. Further, Datta and Mandal [6] generalized the idea of preliminary testing to the uncertain random effects in the Fay-Herriot model, which is described as

$$y_i = \theta_i + \varepsilon_i, \quad \theta_i = \mathbf{x}_i^\top \boldsymbol{\beta} + u_i v_i, \quad i = 1, \dots, m,$$

where  $\varepsilon_i \sim \mathcal{N}(0, D_i)$  for known  $D_i$ ,  $v_i \sim \mathcal{N}(0, A)$  and  $\Pr(u_i = 1) = p = 1 - \Pr(u_i = 0)$ . In Datta and Mandal [6], the term  $u_i v_i$  is called the “uncertain random effect” since the density of  $u_i v_i$  is expressed as a mixture of  $\mathcal{N}(0, A)$  and the one-point distribution on 0. The mixture expression of the distribution of random effects can control the extent of random effects and flexible prediction can be achieved. Actually, the resulting estimator (predictor) of  $\theta_i$  is expressed as the linear combination of the direct estimator  $y_i$  and the regression estimator  $\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}$ . The weight depends on the squared residuals  $(y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}})^2$  while the weight in the resulting estimator from the traditional Fay-Herriot model does not take the residuals into account. In Datta and Mandal [6], the Bayesian method was implemented for inferences of the small area parameters  $\theta_i$ ’s as well as the model parameters by setting the proper prior distributions for  $p$  and  $A$ , namely  $p \sim \text{Beta}(a_1, a_2)$  and  $A \sim \text{IG}(a_3, a_4)$  for known (user specified)  $a_i$ ,  $i = 1, 2, 3, 4$ , and the improper uniform prior for  $\boldsymbol{\beta}$ , where  $\text{Beta}(a_1, a_2)$  and  $\text{IG}(a_3, a_4)$  denote the beta and inverse gamma distributions, respectively. It was shown that the resulting posterior distributions of all the parameters are proper under some conditions. However, Datta and Mandal [6] focused on the Fay-Herriot model, and their method could be restrictive in real applications. Moreover, they used the proper (informative) prior distribution for both  $p$  and  $A$ , and the result could be affected by the choice of hyperparameters.

In this paper, we treat not only the uncertain random effects in more general small area models like the NER model, but also non-informative prior distributions for model parameters. The NER model has been used in various applications including small area estimation, biological experiments and econometric

analysis. The NER model is described as

$$y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + v_i + \varepsilon_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, m,$$

where  $\varepsilon_{ij}$  is the sampling error associated with  $y_{ij}$  and  $v_i$  is a random effect in the  $i$ th area. It is usually assumed that  $\varepsilon_{ij}$  and  $v_i$  are mutually independent and distributed as  $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$  and  $v_i \sim \mathcal{N}(0, \tau^2)$ , respectively. The main purpose of the NER model is to predict (estimate) the quantity of linear combinations of  $\boldsymbol{\beta}$  and  $v_i$ , namely  $\mu_i = \mathbf{c}_i^\top \boldsymbol{\beta} + v_i$  for some known vector  $\mathbf{c}_i$ . For a decade, there has been criticism that the assumption of the NER model is not necessarily satisfied in real applications and several extensions of the NER model have been proposed in order to adapt to real data sets. For example, Jiang and Nguyen [10], Kubokawa, Sugawara, Ghosh and Chaudhuri [11] and Sugawara and Kubokawa [6] proposed heteroscedastic nested error regression models in which the variance components  $\tau^2$  and  $\sigma^2$  are not constant over the areas. Also, Ghosh, Sinha and Kim [9], Arima, Datta and Liseo [1] and Torabi [16] introduced extended models with measurement errors in covariates. However, the problem of uncertainty of random effects, to our knowledge, has not been considered so far in the context.

In this article, we suggest the use of the uncertain random effect in the NER model and propose the uncertain nested error regression (UNER) model by adopting the structure

$$v_i | u_i \sim \mathcal{N}(0, u_i \tau^2) \quad \text{with} \quad \Pr(u_i = 1) = p.$$

For the prior distribution of  $\tau^2$ , the variance of random effects, we use the prior distribution depending on  $u_i$ 's, which is defined as

$$\pi(\tau^2 | z > a) \propto \tau^{-1}, \quad \pi(\tau^2 | z \leq a) \propto \pi_*(\tau^2),$$

for some  $a > 0$ , where  $z = \sum_{i=1}^m u_i$  and  $\pi_*(\tau^2)$  is some proper density, so that the prior distribution of  $\tau^2$  is more non-informative than the proper prior such as an inverse gamma distribution as used in Datta and Mandal [6]. For the other parameters  $\boldsymbol{\beta}$ ,  $\sigma^2$  and  $p$ , we also assign the non-informative prior as  $\pi(\boldsymbol{\beta}, \sigma^2, p) \propto p^{-1/2} (1-p)^{-1/2} \sigma^{-1}$ . Hence, our Bayesian procedure is objective. We also apply the NER model in the framework of the finite population to predict the true finite population mean based on the partially observed data in each population.

This article is organized as follows. In Section 2, we describe the details of the UNER model and provide the Bayesian estimation method as well as the main theorem regarding the propriety of the posterior distribution and the finiteness of posterior variances. The prediction problem of finite population means using UNER is also discussed. In Section 3, we compare the UNER model with the NER model through simulation and empirical studies. Concluding remarks are given in Section 4 and the technical proof is given in the Appendix.

## 2. Uncertain Nested Error Regression Models

### 2.1. Model settings and Bayes estimator

We consider the following uncertain nested error regression (UNER) model

$$\begin{aligned} y_{ij} &= \mathbf{x}_{ij}^\top \boldsymbol{\beta} + v_i + \varepsilon_{ij}, \quad j = 1, \dots, n_i, \\ v_i | (u_i = 1) &\sim \mathcal{N}(0, \tau^2), \quad v_i | (u_i = 0) \sim \delta_0(v_i), \quad i = 1, \dots, m, \end{aligned} \quad (2.1)$$

independently for  $i$  with  $\Pr(u_i = 1) = 1 - \Pr(u_i = 0) = p$ , where  $\mathbf{x}_{ij}$  is a  $q$ -dimensional vector of covariates,  $\boldsymbol{\beta}$  is a  $q$ -dimensional vector of regression coefficients,  $\delta_0(\cdot)$  denotes the Dirac measure on 0, and  $\varepsilon_{ij}$ 's are independently and identically distributed as  $\mathcal{N}(0, \sigma^2)$ . The marginal density function of  $v_i$  is given by

$$f(v) = \frac{p}{\sqrt{2\pi}\tau} \exp\left(-\frac{v^2}{2\tau^2}\right) + (1-p)I(v=0),$$

which is a mixture of the normal distribution  $\mathcal{N}(0, \tau^2)$  and the point mass on 0. Thus the model parameters are regression coefficients  $\boldsymbol{\beta}$ , the variance components  $\sigma^2$  and  $\tau^2$ , and the mixture ratio  $p$ . Let  $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^\top$  be the observed vector in the  $i$ th area. Then the variance of  $\mathbf{y}_i$  is  $\text{Var}(\mathbf{y}_i) = \sigma^2 \mathbf{I}_{n_i} + p\tau^2 \mathbf{J}_{n_i}$  for  $\mathbf{J}_{n_i} = \mathbf{1}_{n_i} \mathbf{1}_{n_i}^\top$ . If the prior probability  $p$  of  $u_i = 1$  is 0, it follows that  $\text{Var}(\mathbf{y}_i) = \sigma^2 \mathbf{I}_{n_i}$ , and the observations in the  $i$ th area are mutually independent. The parameter which we want to estimate (predict) is  $\mu_i = \mathbf{c}_i^\top \boldsymbol{\beta} + v_i$  for a known vector  $\mathbf{c}_i$ . The typical choice of  $\mathbf{c}_i$  is  $\bar{\mathbf{x}}_i = n_i^{-1} \sum_{j=1}^{n_i} \mathbf{x}_{ij}$  in which  $\mu_i$  corresponds to the mean of the  $i$ th area.

The posterior distribution of  $\mu_i$  given  $u_i$  and  $\mathbf{y}_i$  is

$$\mu_i | u_i, \mathbf{y}_i \sim N\left(\mathbf{c}_i^\top \boldsymbol{\beta} + \frac{n_i \tau^2 I(u_i = 1)}{\sigma^2 + n_i \tau^2} (\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta}), \frac{I(u_i = 1) \sigma^2 \tau^2}{\sigma^2 + n_i \tau^2}\right),$$

where  $\bar{y}_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$ , the sample mean of  $y_{ij}$  in the  $i$ th area. Thus the posterior distribution of  $\mu_i$  given  $\mathbf{y}_i$  is a mixture of the normal distribution and one point mass on  $\mathbf{c}_i^\top \boldsymbol{\beta}$ . The resulting Bayes estimator  $\tilde{\mu}_i$  of  $\mu_i$  is

$$\begin{aligned} \tilde{\mu}_i &= E[\mu_i | \mathbf{y}_i] = \tilde{p}_i \left\{ \mathbf{c}_i^\top \boldsymbol{\beta} + \frac{n_i \tau^2}{\sigma^2 + n_i \tau^2} (\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta}) \right\} + (1 - \tilde{p}_i) \mathbf{c}_i^\top \boldsymbol{\beta} \\ &= \mathbf{c}_i^\top \boldsymbol{\beta} + \frac{n_i \tau^2 \tilde{p}_i}{\sigma^2 + n_i \tau^2} (\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta}), \end{aligned}$$

where  $\tilde{p}_i$  is the posterior probability of  $u_i = 1$  given by

$$\begin{aligned} \tilde{p}_i &= \Pr(u_i = 1 | \mathbf{y}_i) \\ &= p \left\{ p + (1-p) \sqrt{\frac{\sigma^2 + n_i \tau^2}{\sigma^2}} \exp\left(-\frac{n_i^2 \tau^2}{2\sigma^2(\sigma^2 + n_i \tau^2)} (\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta})^2\right) \right\}^{-1}. \end{aligned} \quad (2.2)$$

We note that  $\tilde{p}_i$  increases in  $p$  and  $(\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta})^2$ . Thus, if  $\mathbf{x}_{ij}$  is a good covariate to explain  $y_{ij}$  in the  $i$ th area, the squared residual  $(\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta})^2$  is expected to be

small, and the posterior probability  $\tilde{p}_i$  is small as well. The posterior probability  $\tilde{p}_i$  is 1 when  $p = 1$  and  $\tilde{p}_i$  converges to 1 as  $(\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta})^2$  goes to infinity.

Moreover, the posterior variance of  $\mu_i$  is expressed as

$$\begin{aligned} V_i(\mathbf{y}_i) &\equiv \text{Var}(\mu_i | \mathbf{y}_i) = \text{Var}(v_i | \mathbf{y}_i) \\ &= \frac{n_i^2 \tau^4}{(\sigma^2 + n_i \tau^2)^2} (\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta})^2 \tilde{p}_i (1 - \tilde{p}_i) + \frac{\sigma^2 \tau^2 \tilde{p}_i}{\sigma^2 + n_i \tau^2}. \end{aligned} \quad (2.3)$$

It is interesting to point out that the posterior variance of  $\mu_i$ , in this case, depends on observation  $\mathbf{y}_i$  through the squared residual  $(\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta})^2$  and the posterior probability  $\tilde{p}_i$ , while the posterior variance of the random effect in the usual nested error regression model is given by  $\sigma^2 \tau^2 (\sigma^2 + n_i \tau^2)^{-1}$ , which does not depend on observation  $\mathbf{y}_i$ . This means that the uncertain random effect enables us to take the distance between sample mean  $\bar{y}_i$  and synthetic estimator  $\bar{\mathbf{x}}_i^\top \boldsymbol{\beta}$  into the posterior variability of the interesting parameter  $\mu_i$ .

## 2.2. Bayesian implementation and posterior distribution

Since the marginal likelihood function of the model parameters  $\boldsymbol{\beta}, \sigma^2, \tau^2$  and  $p$  is rather complex, we consider objective Bayesian inference for the model parameters as well as the random effect  $v_i$ . To this end, we rewrite the model (2.1) as

$$\begin{aligned} y_{ij} | v_i, \boldsymbol{\beta}, \sigma^2 &\sim \mathcal{N}(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + v_i, \sigma^2), \quad j = 1, \dots, n_i, \quad i = 1, \dots, m \\ v_i | u_i, \tau^2 &\sim \mathcal{N}(0, u_i \tau^2), \quad u_i | p \sim \text{Ber}(p), \quad i = 1, \dots, m \end{aligned} \quad (2.4)$$

independently for  $i$ , where  $\text{Ber}(p)$  denotes the Bernoulli distribution. For implementation of full Bayesian inference, we need to set prior distributions on the model parameters. To keep objectivity of inferences, we use the uniform prior distribution on  $\boldsymbol{\beta}$  and the Jeffreys prior distributions on  $\sigma^2$  and  $p$ . On the other hand, the prior distribution of  $\tau^2$  should depend on  $z = \sum_{i=1}^m u_i$ , since  $\tau^2$  cannot be identified for a small value of  $z$ . Thus, for the model parameters, we use the prior distributions

$$\pi(\boldsymbol{\beta}, \sigma^2, p) = p^{-1/2} (1-p)^{-1/2} \sigma^{-1}, \quad \pi(\tau^2 | z) \propto \begin{cases} \tau^{-1} & (z > a) \\ \pi_*(\tau^2) & (z \leq a) \end{cases} \quad (2.5)$$

where  $\pi_*(\tau^2) = (\tau^2)^{-b_1-1} \exp(-b_2/\tau^2)$  for known constants  $b_1 > 3$  and  $b_2 > 0$ . The value of  $a$  is chosen by the user, and this point will be discussed later. It is noted that the prior distribution on  $p$  is proper, but the priors on  $\boldsymbol{\beta}, \sigma^2$  and  $\tau^2$  are improper, so that the posterior propriety is not always guaranteed. In Theorem 2.1, we show that the posterior distribution for the model parameters is proper under mild conditions.

We now describe the posterior distribution and investigate its properties. The set of all observed data is denoted by  $D = \{\mathbf{y}_i, \mathbf{X}_i\}_{i=1, \dots, m}$  for  $\mathbf{X}_i =$

$(\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i})$ . From the model (2.4) with prior setup (2.5), the posterior density of parameters  $(\mathbf{v}, \mathbf{u}, \boldsymbol{\beta}, \sigma^2, \tau^2, p)$  for  $\mathbf{v} = (v_1, \dots, v_m)^\top$  and  $\mathbf{u} = (u_1, \dots, u_m)^\top$  is given by

$$\begin{aligned} & \pi(\mathbf{v}, \mathbf{u}, \boldsymbol{\beta}, \sigma^2, \tau^2, p|D) \\ & \propto (\sigma^2)^{-(N+1)/2} (\tau^2)^{-(z+I(z>a))/2-(b_1+1)I(z\leq a)} p^{z-1/2} (1-p)^{m-z-1/2} \\ & \quad \times \prod_{i=1}^m \left[ \exp \left( -\frac{\sum_{j=1}^{n_i} (y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta} - v_i)^2}{2\sigma^2} - \frac{u_i v_i^2}{2\tau^2} \right) \delta_0(v_i)^{1-u_i} \right] \\ & \quad \times \exp \left\{ -\frac{b_2}{\tau^2} I(z \leq a) \right\}. \end{aligned} \quad (2.6)$$

Now, we state our main result about the posterior propriety and the existence of posterior variances .

**Theorem 2.1.** (a) *The marginal posterior density  $\pi(\boldsymbol{\beta}, \sigma^2, \tau^2, p|D)$  is proper if  $N > q + 2$  and  $m > a \geq 1$*   
(b) *The model parameters  $\boldsymbol{\beta}, \sigma^2, \tau^2$  and  $p$  have finite posterior variances if  $N > q + 6$  and  $m > a \geq 5$ .*

Remember that  $q$  is the dimension of the vector of regression coefficients  $\boldsymbol{\beta}$ , and  $a$  is the tuning parameter of the prior for  $\tau^2$ . Part (a) in Theorem 2.1 says that the marginal posterior densities of the small area means are proper and part (b) provides a sufficient condition for obtaining finite measures of uncertainty for the model parameters. We note that the conditions in Theorem 2.1 are similar to the conditions given in Arima, et al. [1] and Datta and Mandal [6]. The proof of Theorem 2.1 is presented in the Appendix.

Since the posterior distribution in (2.6) cannot be obtained in a closed form, we rely on the Markov chain Monte Carlo technique, in particular the Gibbs sampler, in order to draw samples from the posterior distribution. This requires generating samples from the full conditional distributions for each of  $(\mathbf{v}, \mathbf{u}, \boldsymbol{\beta}, \sigma^2, \tau^2, p)$  given the remaining parameters and the data  $D$ . Fortunately, the full conditional distributions are described as familiar distributions allowing us to easily implement the Gibbs sampling. The full conditional distributions are given by

$$\begin{aligned} v_i | u_i, \boldsymbol{\beta}, \sigma^2, \tau^2, D & \sim N \left( \frac{n_i \tau^2 I(u_i = 1)}{\sigma^2 + n_i \tau^2} (\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta}), \frac{\sigma^2 \tau^2 I(u_i = 1)}{\sigma^2 + n_i \tau^2} \right), \quad i = 1, \dots, m, \\ u_i | \boldsymbol{\beta}, \sigma^2, \tau^2, p, D & \sim \text{Ber}(\tilde{p}_i), \quad i = 1, \dots, m, \quad p | \mathbf{u}, D \sim \text{Beta} \left( z + \frac{1}{2}, m - z + \frac{1}{2} \right), \\ \boldsymbol{\beta} | \mathbf{u}, \sigma^2, \tau^2, D & \sim N_p((\mathbf{X}^\top \boldsymbol{\Sigma}_u^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Sigma}_u^{-1} \mathbf{y}, (\mathbf{X}^\top \boldsymbol{\Sigma}_u^{-1} \mathbf{X})^{-1}), \\ \tau^2 | \mathbf{u}, \mathbf{v}, D & \sim IG \left( \frac{1}{2} (z - I(z > a)) + b_1 I(z \leq a), \frac{1}{2} \sum_{i=1}^m u_i v_i^2 + b_2 I(z \leq a) \right), \\ \sigma^2 | \mathbf{v}, \boldsymbol{\beta}, D & \sim IG \left( \frac{1}{2} (N - 1), \frac{1}{2} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta} - \mathbf{Z} \mathbf{v})^\top (\mathbf{y} - \mathbf{X} \boldsymbol{\beta} - \mathbf{Z} \mathbf{v}) \right), \end{aligned} \quad (2.7)$$

where  $z = \sum_{i=1}^m u_i$ ,  $\Sigma_u = \text{diag}(\Sigma_{1u}, \dots, \Sigma_{mu})$  with  $\Sigma_{iu} = \sigma^2 \mathbf{I}_{n_i} + u_i \tau^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^\top$ ,  $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_m^\top)^\top$ ,  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_m)$ , and  $\tilde{p}_i$  is given in (2.2). Using these expressions of full conditional distributions, we can easily draw posterior samples of all the variances and parameters to make inferences, such as point estimation, prediction intervals and standard errors, for  $\mu_i = \mathbf{c}_i^\top \boldsymbol{\beta} + v_i$ .

In the closing of this section, we discuss the choices of  $a, b_1$  and  $b_2$  in the posterior distribution of  $\tau^2$ . We remember that the prior distribution of  $\tau^2$  is non-informative and improper when  $z > a$  and informative and proper when  $z \leq a$ . Taking it into account, we should select a value of  $a$  as small as possible. Hence, it follows from Theorem 2.1 that  $a = 5$  is the most reasonable choice. On the other hand, as discussed in Datta and Mandal [6], a reasonable choice is  $b_1 = V + 2$  and  $b_2 = V(V + 1)$  such that  $E[\tau^2 | z \leq a] = V$  and  $\text{Var}(\tau^2 | z \leq a) = V^2$ , where  $V$  is the estimated sampling variance given by

$$V = \frac{1}{N - m - q} \sum_{i=1}^m \sum_{j=1}^{n_i} \{y_{ij} - \bar{y}_i - (\mathbf{x}_{ij} - \bar{\mathbf{x}})^\top \hat{\boldsymbol{\beta}}_{\text{OLS}}\}^2.$$

Here,  $\hat{\boldsymbol{\beta}}_{\text{OLS}}$  is the ordinary least squared estimator of  $\boldsymbol{\beta}$ . It should be noted that  $V$  satisfies  $E[V] = \sigma^2$ .

### 2.3. Prediction in finite populations

Here, we consider the problem of predicting the means in finite populations. Assume that there exist  $m$  finite populations and the  $i$ th population consists of  $N_i$  pairs of data  $(Y_{ij}, \mathbf{x}_{ij})$ ,  $j = 1, \dots, N_i$ . It is supposed that  $n_i (< N_i)$  observations are sampled from the  $i$ th population. What we want to predict is the mean of the  $i$ th finite population  $\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} Y_{ij}$ . Assume also that the mean vector of covariates  $\bar{\mathbf{X}}_i = N_i^{-1} \sum_{j=1}^{N_i} \mathbf{x}_{ij}$  is available, which is often encountered in real application (Battese, et al. [2]). Let  $s_i$  and  $r_i$  be collections of indices of sampled and non-sampled observations in the  $i$ th area, respectively, so that  $s_i$  and  $r_i$  satisfy  $s_i \cap r_i = \emptyset$  and  $s_i \cup r_i = \{1, \dots, N_i\}$ . Without loss of generality, we assume that  $s_i = \{1, \dots, n_i\}$  and  $r_i = \{n_i + 1, \dots, N_i\}$ . The Bayes estimator of  $\bar{Y}_i$  under quadratic loss is given by

$$E[\bar{Y}_i | \mathbf{y}_i] = \frac{1}{N_i} \left\{ n_i \bar{y}_{i(s)} + (N_i - n_i) E[\bar{Y}_{i(r)} | \mathbf{y}_i] \right\},$$

where

$$\bar{y}_{i(s)} = n_i^{-1} \sum_{j \in s_i} y_{ij}, \quad \bar{Y}_{i(r)} = (N_i - n_i)^{-1} \sum_{j \in r_i} Y_{ij}.$$

For evaluating the conditional mean  $E[\bar{Y}_{i(r)} | \mathbf{y}_i]$ , we assume that  $Y_{ij}$  is expressed as

$$Y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + v_i + \varepsilon_{ij}, \quad j \in r_i,$$

that is, the non-sampled observations have the same data generating structure as the sampled ones. Then the unobserved mean  $\bar{Y}_{i(r)}$  is expressed as

$$\bar{Y}_{i(r)} = \bar{\mathbf{x}}_{i(r)}^\top \boldsymbol{\beta} + v_i + \bar{\varepsilon}_{i(r)},$$

where  $\bar{\varepsilon}_{i(r)} = (N_i - n_i)^{-1} \sum_{j \in r_i} \varepsilon_{ij}$ . Thus the conditional distribution of  $\bar{Y}_{i(r)}$  given  $\mathbf{y}_i$  and  $u_i$  is

$$\bar{Y}_{i(r)} | \mathbf{y}_i, u_i \sim N \left( \bar{\mathbf{x}}_{i(r)}^\top \boldsymbol{\beta} + \frac{I(u_i = 1) n_i \tau^2}{\sigma^2 + n_i \tau^2} (\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta}), \frac{I(u_i = 1) \sigma^2 \tau^2}{\sigma^2 + n_i \tau^2} + \frac{\sigma^2}{N_i - n_i} \right), \quad (2.8)$$

which yields the predictive density of  $\bar{Y}_{i(r)}$  given by

$$\begin{aligned} \bar{Y}_{i(r)} | \mathbf{y}_i \sim & \tilde{p}_i N \left( \bar{\mathbf{x}}_{i(r)}^\top \boldsymbol{\beta} + \frac{n_i \tau^2}{\sigma^2 + n_i \tau^2} (\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta}), \frac{\sigma^2 \tau^2}{\sigma^2 + n_i \tau^2} + \frac{\sigma^2}{N_i - n_i} \right) \\ & + (1 - \tilde{p}_i) N \left( \bar{\mathbf{x}}_{i(r)}^\top \boldsymbol{\beta}, \frac{\sigma^2}{N_i - n_i} \right), \end{aligned}$$

where  $\tilde{p}_i$  is the posterior probability of  $u_i = 1$  given in (2.2). Thus the conditional distribution of the non-sampled data is a mixture of the two normal distributions of the predictive density, with and without random effect. Moreover, the conditional variance  $\bar{Y}_{i(r)}$  given  $\mathbf{y}_i$  is calculated as  $V_i(\mathbf{y}_i) + (N_i - n_i)^{-1} \sigma^2$ , where  $V_i(\mathbf{y}_i)$  is the posterior variance of  $v_i$  given in (2.3). It is noted that, when the true mean vector of the explanatory variables  $\bar{\mathbf{X}}_i$  is available in each area, the value of  $\bar{\mathbf{x}}_{i(r)}$  is easily obtained by

$$\bar{\mathbf{x}}_{i(r)} = (N_i - n_i)^{-1} (N_i \bar{\mathbf{X}}_i - n_i \bar{\mathbf{x}}_i).$$

To implement the prediction in the finite population model, we regard  $\bar{Y}_{i(r)}$  as latent variables and add the sampling step from (2.8) to the Gibbs sampling given in (2.7).

### 3. Numerical Studies

#### 3.1. Model based simulations

In this simulation study, we compared the UNER model with the conventional NER model in terms of the quality of the estimates. In applying the NER model, we used the Jeffreys prior on  $(\boldsymbol{\beta}, \tau^2, \sigma^2)$ , namely  $\pi(\boldsymbol{\beta}, \tau^2, \sigma^2) = \tau^{-1} \sigma^{-1}$ , where it is well-known that the resulting posterior distribution is proper (Berger [3]). The full conditional posterior distributions are given by

$$\begin{aligned} v_i | \boldsymbol{\beta}, \sigma^2, \tau^2, D & \sim N \left( \frac{n_i \tau^2}{\sigma^2 + n_i \tau^2} (\bar{y}_i - \bar{\mathbf{x}}_i^\top \boldsymbol{\beta}), \frac{\sigma^2 \tau^2}{\sigma^2 + n_i \tau^2} \right), \quad i = 1, \dots, m \\ \boldsymbol{\beta} | \tau^2 \sigma^2, D & \sim N_p((\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}, (\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1}), \\ \tau^2 | \mathbf{v}, D & \sim IG \left( \frac{1}{2} (m - 1), \frac{1}{2} \sum_{i=1}^m v_i^2 \right), \\ \sigma^2 | \mathbf{v}, \boldsymbol{\beta}, D & \sim IG \left( \frac{1}{2} (N - 1), \frac{1}{2} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta} - \mathbf{Z} \mathbf{v})^\top (\mathbf{y} - \mathbf{X} \boldsymbol{\beta} - \mathbf{Z} \mathbf{v}) \right), \end{aligned} \quad (3.1)$$

where  $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m)$  with  $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}_{n_i} + \tau^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^\top$ . We considered the following data generating process:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + \varepsilon_{ij}, \quad j = 1, \dots, n, \quad i = 1, \dots, m,$$



where  $\varepsilon_{ij} \sim \mathcal{N}(0, 1)$ ,  $\beta_0 = 1$ ,  $\beta_1 = 0.5$ , and  $x_{ij}$ 's were generated from the uniform distribution on  $(1, 2)$  and fixed through simulation runs. The four combinations of  $(n, m)$  were considered as  $(n, m) = (5, 20), (5, 40), (10, 20), (10, 40)$ . For the true distributions of  $v_i$ , we considered the following four scenarios for each choice of  $(n, m)$ .

$$\begin{aligned} \text{S1} : v_i &\sim \mathcal{N}(0, (0.7)^2), & \text{S2} : v_i &\sim 0.3\delta_0(v_i) + 0.7\mathcal{N}(0, (0.7)^2), \\ \text{S3} : v_i &\sim 0.3\delta_0(v_i) + 0.7\mathcal{L}(0, (0.7)^2), & \text{S4} : v_i &\sim 0.3\delta_0(v_i) + 0.7t_6(0, (0.7)^2), \end{aligned}$$

where  $t_6(a, b)$  and  $\mathcal{L}(a, b)$  denote the scaled  $t$ -distribution with 6 degrees of freedom with mean  $a$  and variance  $b$  and the scaled Laplace distribution with mean  $a$  and variance  $b$ , respectively. Hence, UNER is misspecified in scenarios S3 and S4, and overspecified in scenario S1.

Based on  $R = 1,000$  simulation runs, we computed the mean squared errors (MSE), absolute bias of  $\hat{\mu}_i$ , and empirical coverage probability of the 95% credible interval of  $\mu_i$ , which are respectively defined as

$$\begin{aligned} \text{MSE} &= \frac{1}{mR} \sum_{r=1}^R \sum_{i=1}^m (\hat{\mu}_i^{(r)} - \mu_i^{(r)})^2, & \text{Bias} &= \frac{1}{mR} \sum_{r=1}^R \sum_{i=1}^m |\hat{\mu}_i^{(r)} - \mu_i^{(r)}| \\ \text{CP} &= \frac{1}{mR} \sum_{r=1}^R \sum_{i=1}^m I(\mu_i^{(r)} \in \text{CI}_i^{(r)}) \times 100, \end{aligned}$$

where  $\hat{\mu}_i^{(r)}$ ,  $\mu_i^{(r)}$  and  $\text{CI}_i^{(r)}$  are the posterior mean, the true value, and the 95% credible intervals, respectively, of  $\mu_i$  in the  $r$ th simulation runs. In each iteration of the simulation run, we used 5,000 posterior samples after 1,000 initial iterations for both UNER and NER. The results are given in Table 1. In scenario S1, both the MSE and absolute bias of UNER are larger than those of NER since UNER is overspecified. However, as the number of  $n$  and  $m$  get large, the difference of these values gets small. For the other scenarios, we can observe that UNER clearly performs better than NER in terms of MSE and absolute bias, and the differences get larger as  $n$  and  $m$  get larger. Finally, it is observed that the coverage probability of credible intervals are similar in UNER and NER. Hence, we can conclude that UNER is expected to be a useful tool when  $m$  and  $n$  are moderate or large.

### 3.2. Application to PLP data in Japan

This example, primarily for illustration, used the UNER model (2.1) and data from the posted land price data along the Keikyu train line in 2001. This train line connects the suburbs in Kanagawa prefecture to the Tokyo metropolitan area. Those who live in the suburbs in Kanagawa prefecture take this line to work in Tokyo every weekday. Thus, it is expected that the land price depends on the distance from Tokyo. The posted land price data are available for 52 stations on the Keikyu train line, and we consider each station as a small area, namely,  $m = 52$ . For the  $i$ th station, data of  $n_i$  land spots are available, where  $n_i$  varies around 4 and ranges from 1 to 11.

Table 1: Simulated MSE, Bias and Coverage Probabilities (CP) of UNER and NER in Different Scenarios.

$(n, m)$	Scenario	UNER			NER		
		MSE	Bias	CP	MSE	Bias	CP
(3, 25)	S1	0.278	0.419	92.3	0.265	0.408	92.3
	S2	0.165	0.308	93.6	0.176	0.320	93.4
	S3	0.156	0.293	93.3	0.166	0.309	93.2
	S4	0.163	0.301	93.9	0.172	0.313	93.8
(3, 50)	S1	0.248	0.396	93.2	0.242	0.388	93.2
	S2	0.126	0.252	94.3	0.136	0.267	94.3
	S3	0.128	0.245	93.6	0.140	0.261	93.6
	S4	0.130	0.258	94.6	0.140	0.272	94.3
(6, 25)	S1	0.160	0.319	93.7	0.154	0.313	93.7
	S2	0.088	0.215	94.1	0.098	0.235	94.1
	S3	0.088	0.217	93.7	0.103	0.239	93.7
	S4	0.094	0.221	93.8	0.104	0.240	93.8
(6, 50)	S1	0.144	0.302	94.3	0.141	0.299	94.3
	S2	0.076	0.206	94.5	0.095	0.229	94.5
	S3	0.071	0.180	94.3	0.091	0.216	94.3
	S4	0.077	0.191	95.1	0.088	0.216	95.1

For  $j = 1, \dots, n_i$ , let  $y_{ij}$  denote the log-transformed value of the posted land price (Yen) per for square meter of the  $j$ th spot,  $T_i$  is the time it takes from the nearby station  $i$  to Tokyo station around 8:30 in the morning,  $D_{ij}$  is the geographical distance from spot  $j$  to station  $i$  and  $FAR_{ij}$  denotes the floor-area ratio, or ratio of the building volume to the area at spot  $j$ . These values of  $T_i$ ,  $D_{ij}$  and  $FAR_{ij}$  are also transformed by the logarithmic function. We applied the UNER model (2.1) described as

$$\begin{aligned} y_{ij} &= \beta_0 + FAR_{ij}\beta_1 + T_i\beta_2 + D_{ij}\beta_3 + v_i + \varepsilon_{ij}, \\ v_i| (u_i = 1) &\sim \mathcal{N}(0, \tau^2), \quad v_i| (u_i = 0) \sim \delta_0(v_i), \end{aligned} \quad (3.2)$$

where  $\varepsilon_{ij}$ 's are independent and identically distributed as  $\mathcal{N}(0, \sigma^2)$ . For comparison, we also applied the conventional NER model to this data set.

In applying the UNER model, we used the prior distribution with  $a = 5$  and  $b_1 = V+2, b_2 = V(V+1)$  for  $V = 0.031$  as discussed in the end of Section 2.2. In both models, we generated 100,000 posterior samples after 10,000 iterations of Gibbs sampling given in (2.7) and (3.1), respectively, and obtained the posterior means as well as the 95% credible intervals of the model parameters, which are given in Table 2. Moreover, based on the posterior samples, we computed the Deviance Information Criterion (DIC) suggested in Spiegelhalter, Best, Carlin and van der Linde [14], which is defined as  $DIC = 2\overline{D(\boldsymbol{\phi})} - D(\overline{\boldsymbol{\phi}})$ , where  $\boldsymbol{\phi}$  is a vector of the unknown model parameters,  $D(\boldsymbol{\phi})$  is  $(-2)$  times the log-

Table 2: Posterior Means and Credible Intervals of the Model Parameters, and DIC.

		$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\sigma^2$	$\tau^2$	$p$	DIC
UNER	95%CI (upper)	15.16	0.24	-0.53	-0.051	0.041	0.071	0.99	512.6
	mean	14.55	0.17	-0.61	-0.091	0.033	0.017	0.54	
	95%CI (lower)	13.88	0.11	-0.69	-0.131	0.026	0.002	0.05	
NER	95%CI (upper)	15.17	0.24	-0.53	-0.050	0.20	0.117	—	703.1
	mean	14.52	0.17	-0.61	-0.089	0.18	0.075	—	
	95%CI (lower)	13.88	0.10	-0.69	-0.132	0.16	0.031	—	

marginal likelihood function, and  $\overline{D(\phi)}$  and  $\overline{\phi}$  denote the posterior means of  $D(\phi)$  and  $\phi$ , respectively. Note that  $\phi = \{\beta, \tau^2, \sigma^2, p\}$  in the UNER model, and  $\phi = \{\beta, \tau^2, \sigma^2\}$  in the NER model, which are given in Table 2 as well. It is revealed from Table 2 that the posterior estimates and credible intervals of regression coefficients  $\beta_1, \dots, \beta_4$  are similar between UNER and NER, and in both models, all the credible intervals of regression coefficients are bounded away from 0. On the other hand, the results of variance components  $\tau^2$  and  $\sigma^2$  are different because of the effect of the parameter  $p$ . In terms of DIC values, the UNER model seems more preferable than the conventional NER model. To see the effects of  $u_i$ , we calculated the posterior probabilities  $\tilde{p}_i$ 's which are illustrated in the left panel in Figure 1. It is revealed that the  $\tilde{p}_i$ 's change dramatically from area to area, and the  $\tilde{p}_i$ 's in most areas are around 0.5 which comes from the posterior mean of  $p = 0.54$  as shown in Table 2.

We next considered estimating the land price of a spot with a floor-area ratio of 100% and a distance of 1000m from the station  $i$ , namely

$$\mu_i = \beta_0 + FAR_0\beta_1 + T_i\beta_2 + D_0\beta_3 + v_i,$$

for  $FAR_0 = \log(100)$  and  $D_0 = \log(1000)$ . Based on the posterior samples, we calculated the point estimates  $\hat{\mu}_i$  and the posterior standard errors. The results are given in the right panel of Figure 1, noting that the mean of the posterior standard errors for all areas in UNER and NER are  $6.5 \times 10^{-2}$  and  $6.8 \times 10^{-2}$ , respectively. We also computed the length of the prediction intervals of  $\mu_i$ , and found that the results are similar to standard errors. It is revealed from Figure 1 that UNER provides better estimates than NER in terms of posterior standard errors in most areas. In some areas, the posterior standard errors of UNER are larger than those of NER when correspondingly the posterior probability  $\tilde{p}_i$  is larger than 0.7 as shown in the left panel of Figure 1. Thus the uncertain random effects may increase the variability of predictors compared to the conventional random effects in the areas where the existence of random effect is strongly supported. This phenomenon was pointed out in Datta and Mandal [6] in the Fay-Herriot model. However, taking the DIC values into account as well, the UNER model works well in this application.

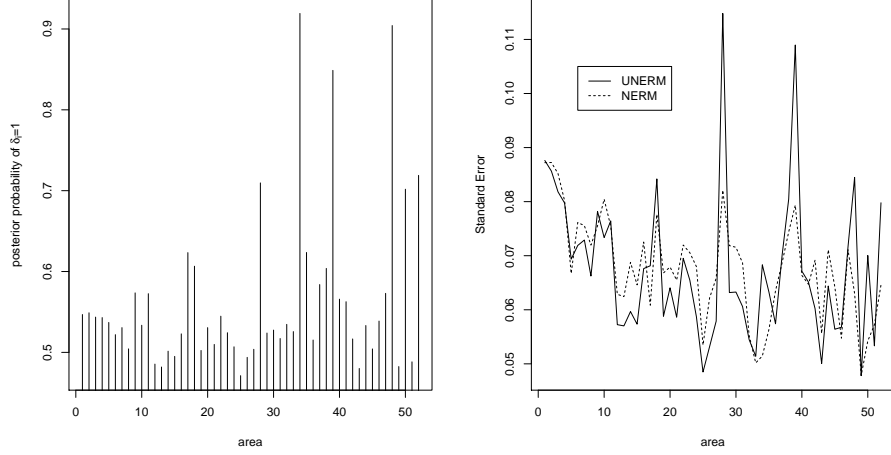


Figure 1: Posterior Probability of  $u_i = 1$  (Left) and Standard Errors of  $\mu_i$  (Right) in Each Area.

### 3.3. Design based simulation

We next investigated the numerical performance of the small area prediction problem in the framework of a finite population. We again used the PLP data in the Kanto region in 2001, which includes Tokyo, Kanagawa, Chiba and Saitama prefectures. Thus the data set includes the PLP data along the Keikyu line used in the previous subsection. The full data set we used is the land price data with covariates ( $T_i$ ,  $D_{ij}$  and  $FAR_{ij}$  as used in the previous study) and each data point has its unique nearest railroad station, which we regard as a small area. For the  $i$ th small area ( $i = 1, \dots, m$ ), there are  $N_i$  land spots. To consider all the observed land price data in each small area in the framework of a finite population, we analyzed only the data which belong to the small areas that have a moderately large number of data points, namely we pick up the area  $i$ 's with  $N_i \geq 20$ . Then the resulting number of finite populations is  $m = 30$ , and the population sizes  $N_i$ 's range from 20 to 45, but most  $N_i$ 's vary around 25. We artificially made the sampled data set and predict each finite population mean of the land price by applying UNER. The sampling scheme is simple random sampling without replacement in each finite population and  $n_i$  data are sampled in the  $i$ th finite population. The sample sizes  $n_i$ 's are decided by some ratio  $0 < \pi < 1$  and  $100\pi$  percent of the data in each population are sampled, that is  $n_i$  is the nearest integer to  $N_i \times \pi$ . We considered four choices for  $\pi$ , namely  $\pi = 0.3, 0.5, 0.7, 0.9$ . In each case, we computed the squared root mean squared errors for estimators of finite population means as

$$\text{SMSE}_i = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\mu}_i^{(r)} - \mu_i)^2},$$

where  $\hat{\mu}_i^{(r)}$  is the estimator of the finite population using UNER or NER, and  $R = 1,000$  in this study. For both UNER and NER, we calculated  $\hat{\mu}_i^{(r)}$  by 5,000 posterior samples after 1,000 iterations using the method discussed in Section 2.3. In the UNER estimation, the same form of the prior distribution as in the previous section was used, namely  $a = 5$ ,  $b_1 = V + 2$  and  $b_2 = V(V + 1)$  for estimated sampling error  $V$ . To compare values of the SMSE for the two models, we then computed the ratio of SMSE given by  $\text{SMSE}_i^{\text{UNER}}/\text{SMSE}_i^{\text{NER}}$ , and provide their values in Figure 2. It is observed from Figure 2 that UNER provides better estimates than NER in some areas, but worse estimates than in several areas for the four cases of  $\pi$ . Moreover, it is also revealed that an improvement of UNER over NER becomes greater as the sampling rate  $\pi$  gets larger.

#### 4. Concluding Remarks

In this article, we have proposed the use of uncertain random effects in the nested error regression model called the UNER model for unit-level data. This can be regarded as an extension of Datta and Mandal [6]. We have used the non-informative priors for all the parameters and proposed Bayesian inferences for the linear combination of fixed effects and random effects as well as the model parameters. We have shown that the posterior distribution is proper and the posterior variances exist under some conditions. Through the simulation study, we have compared the UNER model with the conventional nested error regression (NER) model. It has been revealed that UNER can provide more accurate estimates than that of NER when the underlying distribution of random effects is a mixture of a point mass on the origin and a continuous distribution. Moreover, we have applied UNER together with NER to the PLP data and have found that the UNER model fits better than the NER model in terms of DIC.

#### Acknowledgement

We would like to thank the associate editor and two reviewers for many valuable comments and helpful suggestions which led to an improved version of this paper. The first author was supported in part by Grant-in-Aid for Scientific Research (15J10076) from the Japan Society for the Promotion of Science (JSPS). Research of the second author was supported in part by Grant-in-Aid for Scientific Research (15H01943 and 26330036) from the Japan Society for the Promotion of Science.

#### Appendix A. Proof of Theorem 2.1.

Let  $\pi^*$  be the right side of (2.6). For part (a), we shall show that

$$\sum_{\mathbf{u} \in \{0,1\}^m} \int \pi^*(\mathbf{v}, \mathbf{u}, \boldsymbol{\beta}, \sigma^2, \tau^2, p|D) d\mathbf{v} d\boldsymbol{\beta} d\sigma^2 d\tau^2 dp < \infty,$$

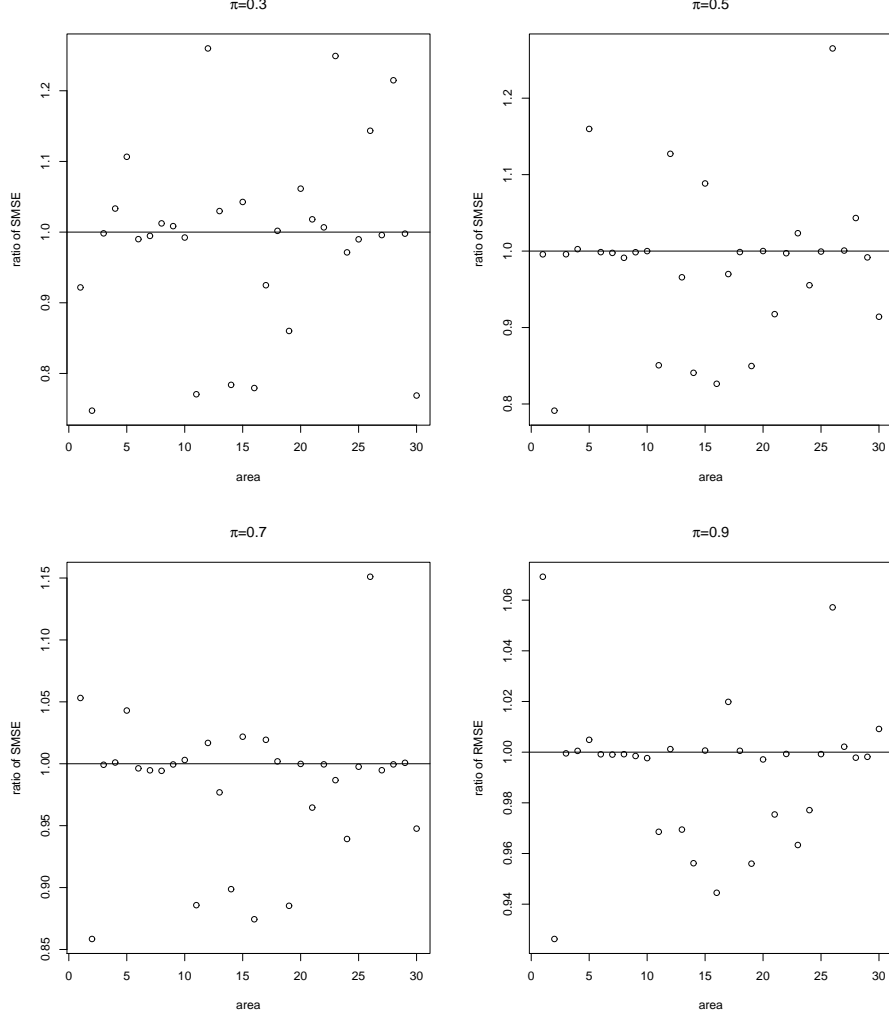


Figure 2: Squared Root Mean Squared Errors of Estimation of Finite Population Mean.

namely the integral for each  $\mathbf{u}$  is finite. We first prove for the case  $\mathbf{u} = (0, \dots, 0)^\top$ . In this case, the integral reduces to

$$\int (\sigma^2)^{-(N+1)/2} (1-p)^{m-1/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta})^2 \right\} d\boldsymbol{\beta} d\sigma^2 dp.$$

It is noted that  $\int p^{-1/2} (1-p)^{m-1/2} dp = B(1/2, m+1/2)$ , where  $B(a, b)$  is a beta function. Then the integral is finite since the posterior distribution of the usual linear regression for the Jeffreys prior is proper if the conditions given in Theorem 2.1 are satisfied.

For the integral in the case  $z \geq 1$ , using  $p^{z-1/2}(1-p)^{m-z-1/2} \leq 1$ , it is sufficient to show that

$$\int \pi_u(\mathbf{v}, \sigma^2, \tau^2, \boldsymbol{\beta}) d\mathbf{v} d\boldsymbol{\beta} d\sigma^2 d\tau^2 < \infty,$$

for

$$\pi_u(\mathbf{v}, \sigma^2, \tau^2, \boldsymbol{\beta}) = \begin{cases} \pi_{u1}(\mathbf{v}, \sigma^2, \tau^2, \boldsymbol{\beta}) & (z > a) \\ \pi_{u2}(\mathbf{v}, \sigma^2, \tau^2, \boldsymbol{\beta}) & (0 < z \leq a) \end{cases} \quad (\text{A.1})$$

where

$$\begin{aligned} \pi_{u1}(\mathbf{v}, \sigma^2, \tau^2, \boldsymbol{\beta}) &= (\sigma^2)^{-(N+1)/2} (\tau^2)^{-(z+1)/2} \prod_{i=1}^m \delta_0(v_i)^{1-u_i} \\ &\quad \times \prod_{i=1}^m \left[ \exp \left( - \frac{\sum_{j=1}^{n_i} (y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta} - v_i)^2}{2\sigma^2} - \frac{u_i v_i^2}{2\tau^2} \right) \right], \end{aligned}$$

and

$$\begin{aligned} \pi_{u2}(\mathbf{v}, \sigma^2, \tau^2, \boldsymbol{\beta}) &= (\sigma^2)^{-(N+1)/2} (\tau^2)^{-z/2} \pi_*(\tau^2) \prod_{i=1}^m \delta_0(v_i)^{1-u_i} \\ &\quad \times \prod_{i=1}^m \left[ \exp \left( - \frac{\sum_{j=1}^{n_i} (y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta} - v_i)^2}{2\sigma^2} - \frac{u_i v_i^2}{2\tau^2} \right) \right]. \end{aligned}$$

To show the integrability of  $\pi_{u1}$  and  $\pi_{u2}$ , we consider the case of  $\mathbf{u}$  with  $\sum_{i=1}^m u_i = k$ . Without loss of generality, we assume that  $u_i = 1$  for  $i = 1, \dots, k$  and  $u_i = 0$  for  $i = k+1, \dots, m$ . Then  $\pi_{u1}(\mathbf{v}, \sigma^2, \tau^2, \boldsymbol{\beta})$  can be rewritten as

$$\begin{aligned} \pi_{u1}(\mathbf{v}, \sigma^2, \tau^2, \boldsymbol{\beta}) &= (\sigma^2)^{-(N+1)/2} (\tau^2)^{-(k+1)/2} \prod_{i=1}^k \left[ \exp \left( - \frac{\sum_{j=1}^{n_i} (y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta} - v_i)^2}{2\sigma^2} - \frac{v_i^2}{2\tau^2} \right) \right] \\ &\quad \times \left[ \prod_{i=k+1}^m \exp \left( - \frac{\sum_{j=1}^{n_i} (y_{ij} - \mathbf{x}_{ij}^\top \boldsymbol{\beta})^2}{2\sigma^2} \right) \delta_0(v_i) \right]. \end{aligned}$$

We define  $N$ -dimensional vector  $\mathbf{s}(\mathbf{v}_*) = (\mathbf{s}_{(1)}(\mathbf{v}_*)^\top, \mathbf{s}_{(2)}^\top)^\top$  as  $\mathbf{s}_{(1)}(\mathbf{v}_*) = ((\mathbf{y}_1 - v_1 \mathbf{1}_{n_1})^\top, \dots, (\mathbf{y}_k - v_k \mathbf{1}_{n_k})^\top)^\top$  and  $\mathbf{s}_{(2)} = (\mathbf{y}_{k+1}^\top, \dots, \mathbf{y}_m^\top)^\top$  for  $\mathbf{v}_* = (v_1, \dots, v_k)^\top$ . Then, if  $N > q$ , we have

$$\begin{aligned} &\int \pi_{u1}(\mathbf{v}, \sigma^2, \tau^2, \boldsymbol{\beta}) d\boldsymbol{\beta} \\ &\propto (\sigma^2)^{-(N-q-1)/2-1} (\tau^2)^{-(k-1)/2-1} \exp \left( - \frac{\mathbf{s}(\mathbf{v}_*)' \mathbf{A} \mathbf{s}(\mathbf{v}_*)}{2\sigma^2} - \frac{\sum_{i=1}^k v_i^2}{2\tau^2} \right), \end{aligned} \quad (\text{A.2})$$

where  $\mathbf{A} = \mathbf{I}_N - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ . The right side is integrable with respect to  $\sigma^2$  and  $\tau^2$  since  $N > q + 1$  and  $k \geq a > 1$ , whereby we obtain

$$\int \pi_{u1}(\mathbf{v}, \sigma^2, \tau^2, \boldsymbol{\beta}) d\boldsymbol{\beta} d\sigma^2 d\tau^2 \propto \pi_{u1}(\mathbf{v}_*) \prod_{i=k+1}^m \delta_0(v_i),$$

where

$$\pi_{u1}(\mathbf{v}_*) = \left\{ \mathbf{s}(\mathbf{v}_*)^\top \mathbf{A} \mathbf{s}(\mathbf{v}_*) \right\}^{-(N-q-1)/2} \left( \mathbf{v}_*^\top \mathbf{v}_* \right)^{-(k-1)/2}.$$

In what follows, we show that  $\pi_{u1}(\mathbf{v}_*)$  is integrable. To this end, we note that

$$\int_{\mathbb{R}^k} \pi_{u1}(\mathbf{v}_*) d\mathbf{v} = \int_{\mathbf{v}_*^\top \mathbf{v}_* \leq 1} \pi_{u1}(\mathbf{v}_*) d\mathbf{v} + \int_{\mathbf{v}_*^\top \mathbf{v}_* \geq 1} \pi_{u1}(\mathbf{v}_*) d\mathbf{v},$$

and we evaluate the two integrals separately. For the first term, since  $\mathbf{A}$  is idempotent and  $\text{rank}(\mathbf{A}) = N - q (> 0)$ , there exists  $c(\mathbf{y}) > 0$  such that  $\mathbf{s}(\mathbf{v}_*)^\top \mathbf{A} \mathbf{s}(\mathbf{v}_*) > c(\mathbf{y})$  for all  $\mathbf{v}_*$ . Then we have

$$\begin{aligned} \int_{\mathbf{v}_*^\top \mathbf{v}_* \leq 1} \pi_{u1}(\mathbf{v}_*) d\mathbf{v} &\leq c^{-(N-q-1)/2} \int_{\mathbf{v}_*^\top \mathbf{v}_* \leq 1} \left( \mathbf{v}_*^\top \mathbf{v}_* \right)^{-(k-1)/2} d\mathbf{v} \\ &= c^{-(N-q-1)/2} V(S^k) \int_0^1 (r^2)^{-(k-1)/2} (r^2)^{(k-1)/2} dr < \infty, \end{aligned}$$

where  $V(S^k)$  is the volume of the unit sphere in  $\mathbb{R}^k$ . For the second term, it follows that

$$\int_{\mathbf{v}_*^\top \mathbf{v}_* \geq 1} \pi_{u1}(\mathbf{v}_*) d\mathbf{v} = \int_{\mathbf{v}_*^\top \mathbf{v}_* \geq 1} \left\{ \mathbf{s}(\mathbf{v}_*)^\top \mathbf{A} \mathbf{s}(\mathbf{v}_*) \right\}^{-(N-q-1)/2} \left( \mathbf{v}_*^\top \mathbf{v}_* \right)^{-(k-1)/2} d\mathbf{v}.$$

Since  $\mathbf{s}(\mathbf{v}_*)^\top \mathbf{A} \mathbf{s}(\mathbf{v}_*)$  is a quadratic function of  $\mathbf{v}_*$ , the integral is finite as far as  $N > q + 2$ . For the integrability of  $\pi_{u2}$ , we carry out integration with respect to  $\boldsymbol{\beta}, \sigma^2$  and  $\tau^2$  to get

$$\int \pi_{u2}(\mathbf{v}, \sigma^2, \tau^2, \boldsymbol{\beta}) d\boldsymbol{\beta} d\sigma^2 d\tau^2 \propto \pi_{u2}(\mathbf{v}_*) \prod_{i=k+1}^m \delta_0(v_i).$$

Since for  $N > q + 1$ ,

$$\begin{aligned} \pi_{u2}(\mathbf{v}_*) &= \left\{ \mathbf{s}(\mathbf{v}_*)^\top \mathbf{A} \mathbf{s}(\mathbf{v}_*) \right\}^{-(N-q-1)/2} \left( \mathbf{v}_*^\top \mathbf{v}_* + 2b_2 \right)^{-k/2-b_1} \\ &\leq c^{-(N-q-1)/2} (2b_2)^{-k/2-b_1}, \end{aligned}$$

it follows that  $\pi_{u2}(\mathbf{v}_*)$  is integrable as far as  $N > q + 1$ . Thus the proof of part (a) is established.

For the proof of part (b), it is sufficient to show that the posterior second moments are finite. Since the statement for  $p$  is clear, we establish the result



for  $\beta, \sigma^2$  and  $\tau^2$ . As in the proof of part (a), we consider the three cases where  $z > a$ ,  $0 < z \leq a$  and  $z = 0$ . By replacing  $N + 1$  in expressions of  $\pi_{u1}, \pi_{u2}$  and  $\pi_{u3}$  with  $N + 5$ , it follows that  $E[(\sigma^2)^2|D] < \infty$  when  $N > q + 6$ .

For  $E[\beta\beta^\top|D]$ , we first note that

$$\begin{aligned} & \int_{\mathbb{R}^q} \beta\beta^\top \exp\left(-\frac{(\mathbf{s}(\mathbf{v}_*) - \mathbf{X}\beta)^\top(\mathbf{s}(\mathbf{v}_*) - \mathbf{X}\beta)}{2\sigma^2}\right) d\beta \\ &= (\sigma^2)^{q/2} |\mathbf{X}^\top \mathbf{X}|^{-1/2} \exp\left(-\frac{\mathbf{s}(\mathbf{v}_*)^\top \mathbf{A} \mathbf{s}(\mathbf{v}_*)}{2\sigma^2}\right) (\mathbf{X}^\top \mathbf{X})^{-1} \\ & \quad \times \left\{ \sigma^2 \mathbf{I}_q + \mathbf{X}^\top \mathbf{s}(\mathbf{v}_*) \mathbf{s}(\mathbf{v}_*)^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \right\} \\ &= (\sigma^2)^{(q+2)/2} h(\mathbf{X}, \mathbf{s}(\mathbf{v}_*), \sigma^2) \\ & \quad + (\sigma^2)^{q/2} h(\mathbf{X}, \mathbf{s}(\mathbf{v}_*), \sigma^2) \mathbf{X}^\top \mathbf{s}(\mathbf{v}_*) \mathbf{s}(\mathbf{v}_*)^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}, \end{aligned}$$

for

$$h(\mathbf{X}, \mathbf{s}(\mathbf{v}_*), \sigma^2) = |\mathbf{X}^\top \mathbf{X}|^{-1/2} \exp\left(-\frac{1}{2\sigma^2} \mathbf{s}(\mathbf{v}_*)^\top \mathbf{A} \mathbf{s}(\mathbf{v}_*)\right) (\mathbf{X}^\top \mathbf{X})^{-1}.$$

Then it follows that

$$\begin{aligned} & \int \beta\beta^\top \pi_{u1}(\mathbf{v}, \beta, \sigma^2, \tau^2) d\mathbf{v} d\beta d\sigma^2 d\tau^2 \\ & \propto \mathbf{I}_q \int_{\mathbb{R}^k} \left\{ \mathbf{s}(\mathbf{v}_*)^\top \mathbf{A} \mathbf{s}(\mathbf{v}_*) \right\}^{-(N-q-3)/2} \left( \mathbf{v}_*^\top \mathbf{v}_* \right)^{-(k-1)/2} d\mathbf{v} \\ & \quad + \int_{\mathbb{R}^k} \mathbf{X}^\top \mathbf{s}(\mathbf{v}_*) \mathbf{s}(\mathbf{v}_*)^\top \mathbf{X} \pi_{u1}(\mathbf{v}_*) d\mathbf{v}. \end{aligned}$$

Since  $\mathbf{v}_* \mathbf{v}_*^\top \leq (\mathbf{v}_*^\top \mathbf{v}_*) \mathbf{I}_q$ , the second term is finite if  $k > 5$  for all  $k \geq a$ , namely  $a > 5$ . The first term is also finite as far as  $N > q + 4$ . For the other cases  $0 < z \leq a$  and  $z = 0$ , we can similarly show that  $\int \beta\beta^\top \pi_{u2}(\mathbf{v}, \beta, \sigma^2, \tau^2) d\mathbf{v} d\beta d\sigma^2 d\tau^2$  and  $\int \beta\beta^\top \pi_{u3}(\mathbf{v}, \beta, \sigma^2, \tau^2) d\mathbf{v} d\beta d\sigma^2 d\tau^2$  are finite under the condition given in Theorem 2.1.

Finally, for  $E[\tau^2|D]$ , it follows that

$$\begin{aligned} & \int (\tau^2)^2 \pi_{u1}(\mathbf{v}, \beta, \sigma^2, \tau^2) d\mathbf{v} d\beta d\sigma^2 d\tau^2 \\ & \propto \int (\sigma^2)^{-(N-p-1)/2-1} (\tau^2)^{-(k-5)/2-1} \\ & \quad \times \exp\left(-\frac{\mathbf{s}(\mathbf{v}_*)^\top \mathbf{A} \mathbf{s}(\mathbf{v}_*)}{2\sigma^2} - \frac{\sum_{i=1}^k v_i^2}{2\tau^2}\right) d\mathbf{v} d\tau^2 d\sigma^2 \\ & \propto \int \left\{ \mathbf{s}(\mathbf{v}_*)^\top \mathbf{A} \mathbf{s}(\mathbf{v}_*) \right\}^{-(N-p-1)/2} \left( \mathbf{v}_*^\top \mathbf{v}_* \right)^{-(k-5)/2} d\mathbf{v} \end{aligned}$$

which is finite as far as  $k > 5$  for all  $k \geq a$ , namely  $a > 5$ . In the cases of  $0 < z \leq a$  and  $z = 0$ , it is integrable if

$$\int_0^\infty (\tau^2)^2 \pi_*(\tau^2) d\tau^2 < \infty,$$

which can be established since  $b_1 > 3$ . Thus we complete the proof of part (b).

## References

## References

- [1] Arima, S., Datta, G. S. and Liseo, B. (2015). Bayesian estimators for small area models when auxiliary information is measured with error. *Scandinavian Journal of Statistics*, **42**, 518-529.
- [2] Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An error component model for prediction of mean crop areas using survey and satellite data. *Journal of the American Statistical Association*, **95**, 28-36.
- [3] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. New York: Springer-Verlag.
- [4] Datta, G. and Ghosh, M. (2012). Small area shrinkage estimation. *Statistical Science*, **27**, 95-114.
- [5] Datta, G., S., Hall, P. and Mandal, A. (2011). Model selection by testing for the presence of small-area effects, and application to area-level data. *Journal of the American Statistical Association*, **106**, 362-374.
- [6] Datta, G., S. and Mandal, A. (2015). Small area estimation with uncertain random effects. *Journal of the American Statistical Association*, **110**, 1735-1744.
- [7] Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269-277.
- [8] Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, **9**, 55-93.
- [9] Ghosh, M., Sinha, K. and Kim, D. (2006). Empirical and hierarchical Bayesian estimation in finite population sampling under structural measurement error models. *Scandinavian Journal of Statistics*, **33**, 591-608.
- [10] Jiang, J. and Nguyen, T. (2012). Small area estimation via heteroscedastic nested-error regression. *Canadian Journal of Statistics*, **40**, 588-603.
- [11] Kubokawa, T., Sugawara, S., Ghosh, M. and Chaudhuri, S. (2015). Prediction in heteroscedastic nested error regression models with random dispersions. *Statistica Sinica*, **26**, 465-492.
- [12] Pfeiffermann, D. (2014). New important developments in small area estimation. *Statistical Science*, **28**, 40-68.

- [13] Rao, J.N.K. and Molina, I. (2015) *Small Area Estimation, 2nd Edition*. Wiley.
- [14] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, B.*, **64**, 583-639.
- [15] Sugasawa, S. and Kubokawa, T. (2016). Heteroscedastic nested error regression models with variance functions. *Statistica Sinica*, to appear.
- [16] Torabi, M. (2012). Small area estimation using survey weights under a nested error linear regression model with structural measurement error. *Journal of Multivariate Analysis*. **109**, 52-60.